Postprint of Food Control Volume 120, February 2021, 107526

DOI: https://doi.org/10.1016/j.foodcont.2020.107526

Viability of near infrared spectroscopy for a rapid analysis of the bioactive compounds in intact cocoa bean husk

Authors: Carolina Hernández-Hernández[a*], Víctor M. Fernández-Cabanás[b], Guillermo Rodríguez-Gutiérrez[c], Alejandra Bermúdez-Oria[c], Ana Morales-Sillero[b]

[a] Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Campo Experimental Huimanguillo. Km 1. Ctra. Federal Huimanguillo- Cárdenas, C.P. 86400. Huimanguillo, Tabasco, México.

[b] Departamento de Ciencias Agroforestales, ETSIA, Universidad de Sevilla, Ctra. Utrera Km 1, 41013, Seville, Spain.

[c]Instituto de la Grasa, Consejo Superior de Investigaciones Científicas (CSIC), Campus Universitario Pablo de Olavide, Edificio 46, Ctra. de Utrera, km. 1 - 41013, Seville, Spain.

Email address:

Ana Morales-Sillero: amorales@us.es

Víctor M. Fernández-Cabanás: victorf@us.es

Guillermo Rodríguez-Gutiérrez: guirogu@cica.es

Alejandra Bermúdez-Oria: aleberori@ig.csic.es.

* Corresponding author: Carolina Hernández Hernández

E-mail address: carolinja89@gmail.com

Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Campo Experimental Huimanguillo. Km 1. Ctra. Federal Huimanguillo- Cárdenas, C.P. 86400. Huimanguillo, Tabasco, México. PHONE +52 1 9171071443

ABSTRACT

The potential of the cocoa bean husk (CBH) as a natural source of bioactive compounds is ever-increasing. In this work, its bioactive compounds and antioxidant activity were analyzed using near infrared spectroscopy in samples of CBH. Beans were harvested and fermented in a Mexican gene bank. Reference data on total sugars, total phenols, phenolic compounds, theobromine, and antioxidant activity were correlated with the intact husk and bean spectra. The Modified Partial Least Square regression method (MPLSR) was used to develop calibrations. Good calibration statistics were obtained for total sugars (r^2=0.90), theobromine (r^2=0.83) and total phenols (r^2=0.81) in data related to the CBH spectra, with a ratio of standard deviation/standard error of cross validation (RPD) of 3.16, 2.39 and 2.28, respectively. Acceptable calibrations for the estimation of bioactive compounds in CBH were obtained for the first time from the spectra of intact grain samples. Industries interested in bioactive compounds from CBH could use this technology as an easy and fast method to predict their contents, while avoiding the inconvenient de-husking process.

Keywords: *Theobroma cacao* L.; total sugars; theobromine; total phenols; antioxidant activity.

1. INTRODUCTION

The cocoa bean husk (CBH) is the main residue of the cocoa industry. Annual worldwide production is estimated at approximately 4,200,000 tones (FAOSTAT, 2018) and expected to increase given the demand for cocoa products. In the cocoa industry, CBH, also named shell, is separated from the cotyledons once the bean is fermented and dried, during or after the pre-roasting process.(Okiyama, Navarro, & Rodrigues, 2017). CBH has been considered a by-product, and traditionally has had limited applications, mainly as animal feed or organic soil fertilizer. In the search for alternatives for its valorization, as for other agro-industrial residues, the study of its composition and possible industrial applications has received more attention in recent years (Okiyama et al., 2017; Panak Balentić et al., 2018).

CBH represents between 12 and 20% of the total bean weight. It has a high fiber content (about 50-60% of its total weight), depending on whether it is roasted or not (Panak Balentić et al., 2018), along with minerals, proteins and all the essential amino acids. The low contents in soluble sugars and fat in unfermented CBH gives it a low calorific value, although the fat has an interesting profile, rich in palmitic, stearic and oleic fatty acids, which highlight its nutritional value. Furthermore, CBH is particularly rich in phenols and methylxanthines, compounds that are stored in the bean cotyledons but diffuse in part into the husk during the fermentation process, where they can accumulate in high concentrations (Arlorio, Coisson, Restani, & Martelli, 2001; Lecumberri, Mateos, et al., 2007). Phenols such as catechin, epicatechin and *p*-hydroxybenzoic acid have been identified (Arlorio et al., 2005; Hernández-Hernández et al., 2018b), as well as theobromine and caffeine methylxanthines (Hernández-Hernández et al., 2018b). This composition in phenols and methylxanthines makes it an interesting source of bioactive compounds, given their antioxidant activity (Martínez et al., 2012). This by-product has been proposed as an inexpensive source of dietary fiber to help reduce calories and cholesterol levels and to control glucose levels in the

blood. (Okiyama et al., 2017). Phenol-rich CBH extracts also have a powerful anti-cariogenic potential as phenols have anti-glucosyltransferase activity (Osawa et. al., 2001). Hartati (2010) attributed health benefits to CBH's theobromine content due to its anti-cancer, diuretic, smooth-muscle relaxant and cardiac stimulant functions.

Therefore, its potential as a natural source of bioactive compounds is moving the interest of researchers and industry toward including CBH extracts as natural additives in food, pharmaceutical and cosmetic products in order to increase their bioactive characteristics. The production of different cocoa extracts has recently been patented (Okiyama et al., 2017; Panak Balentićet al., 2018) .In a previous work, we proposed a totally physical method for the production of a natural extract from CBH on an industrial scale for the first time. This extract is rich in sugars (220 mg g^{-1}), phenols (55 mg g^{-1}) and theobromine (56 mg g^{-1}), and can be used directly, even though the compounds can be easily purified (Hernández-Hernández et al., 2018a).

Not all the raw materials destined for the cocoa industry are of similar interest in terms of their bioactive composition. It depends on their genetic basis, origin and processing (Hernández-Hernández, et al., 2018a; Okiyama et al., 2017). Moreover, analytical methods which are considered appropriate for the identification and quantification of the bioactive compounds in CBH, such as high performance liquid chromatography (Arlorio et al., 2005; Hernández-Hernández et al., 2018a), are tedious and expensive for routine screening purposes. To our knowledge, indirect methods such as near infrared spectroscopy (NIRS) have not yet been explored for the quantification of the bioactive compounds in CBH. This technology is increasingly accepted for the routine analysis of antioxidants in many food, plant and agricultural products, saving analysis time and costs for both industry and research (Cozzolino, 2015). NIRS is based on the rule that the main components of each product, such as water, protein, fat and carbohydrates, exhibit electromagnetic absorption at wavelengths in the range 780–2500 nm. It is a powerful tool for characterizing and classifying foods according to quality standards. Sample preparation is usually quite

simple and numerous parameters can be analyzed at the same time. Current NIR instruments allow fast and low cost measurements and utilize easy-to-use software for building calibration models which relate spectral data with individual chemical components (Alander et al., 2013). Nowadays it is routinely used in different industries at a laboratory level and also at-line, on-line or in-line (Huang et al., 2008).

Regarding CBH, NIRS has been applied for the rapid detection of husk in cocoa powders (Quelal-Vásconez et al., 2019) and for authentication according to geographic origin (Mandrile et al., 2019). The hypothesis of this work is that this technology can also be used in the food industry for the predicting the presence of any interesting bioactive compounds (sugars, phenols and methylxanthines). The content in total sugars includes not only monosaccharides but also potential neutral and acidic oligo and polysaccharides with antioxidant and biological properties like antioxidant fibers, phenolic glycoside modified pectin and prebiotic oligosaccharides previously identified in CBH and other lignocellulosic by-products (Hernández-Hernández et al., 2018b; Lama-Muñoz et al., 2012; Rubio-Senent et al., 2013).The potential for predicting these compounds from the spectra of both unground husk and intact cocoa beans was also evaluated.

2. MATERIALS AND METHODS

2.1. Material

A total of 80 samples of cocoa beans previously fermented and dried to 7% humidity, belonging to 63 different genotypes and harvested in the 2012-2013 (23), 2013-2014 (24) and 2014-2015 (34) seasons were provided by the National Institute of Agricultural and Livestock Forestry Research Germplasm Bank, Mexico from two of its experimental fields (Rosario Izapa and Huimanguillo).

2.3. Physicochemical analyses

2.3.1. Extractions

Two extractions were made for the determination of total sugars, total and individual phenols, as well as theobromine and antioxidant activity, as described by Hernández-Hernández et al. (2018b), in which a methanol:water ratio of 80:20 was used as solvent, adjusting the pH to 3 with 5% HCl and keeping the mixture in a water bath at 70 °C for one hour.

2.3.2. Total sugars

Total sugars were determined using the anthrone method as described by Witham et al. (1971). Absorbance at 630 nm was measured using a spectrophotometer (BIO-RAD iMark Microplate Reader, USA).

2.3.3. Total phenols

Total phenols were determined according to the Folin-Ciocalteu's method as described by Singleton et al. (1998) and expressed as grams of gallic acid equivalents per gram of dried and de-fatted sample. Absorbance at 655 nm was measured using the same spectrophotometer.

2.3.4. Analysis of methylxanthines and individual phenols by High Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD)

Theobromine, catechin, epicatechin and its four derivatives (compounds which come from epicatechin and maintain most of their chemical structure after a chemical or enzymatic reaction), and epigallocatechin were determined as described by Hernández-Hernández et al. (2018b), using a Varian ProStar liquid chromatography system with a C-18 column (Kinetex® Biphenyl 100 Å, 250 mm x 4.6 mm, i.d. 5µm) and a diode array detector (DAD) with automatic Rheodyne injection valves (20 µL loop). All compounds were detected at 280 nm following the methos described in a previous work (Hernández-Hernández et al., 2018a). Calibration curves were constructed for theobromine and (-)-epicatechin at concentrations ranging from 0 to 2 mg/mL ($r^2 \geq 0.99$) and for (+)-catechin at concentrations ranging from 0 to 1 mg/mL (r^2=0.99). The samples were filtered (0.45µm) before injection. All determinations were made in triplicate.

2.3.5. Antioxidant activity

The antioxidant activity of each sample was determined by the 2, 2-diphenyl-1-picrylhydrazil (DPPH) method as described in previous studies (Hernández-Hernández et al., 2018a, 2018b). The result of the activity of each extract was expressed as an EC_{50} (effective concentration at 50% in mg/mL) calculated from a calibration curve by linear regression.

2.4. Statistical analysis

The mean values, range, standard deviation, and coefficient of variation were determined for each parameter. The correlation coefficients between the mean values

were assessed by means of the Pearson's Correlation test. Statgraphics Centurion XVI v. 16.1.15 software was used.

2.5. Scanning of NIR spectra

The reflectance spectra were obtained in a Foss-NIRSystem 6500 SY-II monochromatic device (Foss NIRSystems, Silver Spring, MD). Scanning was carried out from 400 to 2498 nm, every 2 nm (spectral band pass 10 nm ± 1 nm). Intact cocoa bean samples were scanned in a transport module using a rectangular sample cup (Natural Product Sample Cup IH-0331), with dimensions of 4.7 cm in width and 20 cm in length. A smaller cell with dimensions of 6 x 10 cm was used for cocoa husk. The spectra of each sample were obtained as the average of two subsamples.

The WINISI v. 1.5 software. (Infrasoft International, State College, PA) was used to manipulate and process spectral data.

2.6. Development of calibrations and handling of samples with outlier data

To develop calibrations, both the spectra of cocoa husk and the spectra of intact cocoa beans were related to reference analyses obtained for husk samples. Five spectral derivatives were compared to select the best calibrations: 1, 5, 5, 1; 1, 10, 5, 1; 1, 10, 10, 1; 2, 5, 5, 1 and 2, 10, 5, 1. The first digit refers to the number of the derivative, the second is the gap on which the derivative was calculated, the third is the smoothing segment, and the fourth is the second smoothing segment (Marten, Shenk, & Barton, 1989). Calibration equations were developed according to the Modified Partial Least Square Regression method (MPLSR). Dispersion phenomena were corrected by Standard Normal Variate (SNV) and Detrending (DT) mathematical pre-

treatments (Barnes, Dhanoa, & Lister, 1989). A cross-validation method was used to determine the optimum number of PLS terms in the regression models and to avoid overfitting. The validation errors were combined to obtain a standard error cross-validation error (SECV). Two anomalous filters (T and H) were carried out before completing the final calibration, excluding samples with spectral or chemical composition discrepancies. The calibration statistics were: standard error of calibration (SEC), calibration coefficient of determination (R^2), standard error of cross-validation (SECV), coefficient of determination for cross-validation (r^2) and the ratio of SD/SECV (RPD). The best calibration model was selected based on the highest r^2 value and the lowest SECV value. Although the use of external validation is a common tool for the evaluation of NIR predictive models, it should be noted that for fruits and other agricultural products with significant genetic variability, NIR calibrations for one season may not be appropriate for use in the next, as has been reported in several studies (Guthrie, Wedding, & Walsh, 1998; León, Garrido-Varo, & Downey, 2004; Morales-Sillero et al., 2011; Peiris, Dull, Leffler, & Kays, 1998). For this reason, since the samples studied in this work were extremely variable (63 different genotypes with harvests from 3 seasons) and taking into account that the estimation of bioactive principles from the NIR spectra is complex and requires a broad calibration data set, in this feasibility study no external validation was made due to the limited number of samples. Nevertheless, the SECV is the single best estimate of the standard error of prediction (SEP), and is similar to the average SEP obtained from 10 randomly-chosen prediction sets (Shenk and Westerhaus, 1996).

3. RESULTS AND DISCUSSION

3.1. Chemical composition of cocoa bean husk.

The cocoa husk represented around 18% of the bean's weight (minimum value 10%, maximum 31%). The 80 samples of CBH showed a wide range of values for all the compounds analyzed (Table 1), as the coefficients of variation were always greater than 40%. This variation is important in order to develop a good calibration model since a greater range of samples is being represented. The wide variability among the samples was mainly due to their origin, since 76% belonged to different cacao genotypes. Furthermore, the growing environmental conditions of each year in which the bean samples were harvested may have influenced the contents of these compounds in the husk. The mean value for total sugars was 72 mg/g (range 12–337). The concentration in unfermented CBH reported by other authors was lower than 22 mg/g (Martin-Cabrejas et al., 1994). This difference may indicate that soluble sugars are transferred from the cotyledon to the husk during the fermentation process, as described for phenols and theobromine (Hernández-Hernández et al., 2018a, 2018b).The mean value for total phenols was 24 mg/g. Lower contents in these compounds were reported in CBH and varied between 18 and 58 mg/g (Arlorio et al., 2001). For most samples, the CBH phenol contents were lower than those found in their respective cotyledons. For example, total phenols ranged between 15 and 168 mg/g in cotyledons and between 7 and 80 mg/g in the husk for the same samples (Hernández-Hernández, Morales-Sillero et al., 2018).This could be due to the fact that a low proportion of phenols was transferred to the husk during fermentation, or the phenols transferred were linked to the cell wall material (i.e. they formed antioxidant carbohydrates with phenolic glucosides), making their extraction and quantification difficult (Rubio-Senent et al., 2013).Catechin, epicatechin and some of its derivatives, as well as epigallocatechin were quantified. Four isomers of ethyl-linked epicatechin as well as several isomers of epicatechin-ethyl-procyanidin have been identified as metabolites of the microorganisms during the fruit fermentation process (Fayeulle et al., 2018).The mean value for epicatechin was 16 mg/g (range 4–46 mg/g), but for the other phenolic compounds it was lower than 1.5 mg/g. In agreement with Adamson et

al. (1999) and Arlorio et al. (2005), epicatechin was the major phenolic compound in CBH, although they reported contents of just 2.7 mg/g. In relation to methylxanthines, the mean value for theobromine was 19.70 mg/g (7.39-54.59 mg/g). Arlorio et al. (2001) also reported a lower content in this compound in CBH (13 mg/g). As for soluble sugars and phenols, this compound migrated from the cotyledons to the husk during the fermentation process (Hernández-Hernández et al., 2018b; Okiyama et al., 2017).

Concerning antioxidant activity, the mean value was 27 mg/mL (range 6–77 mg/mL). This parameter was estimated from the free radical scavenging activity determined according to the quenching of DPPH radicals and reported as effective concentration at 50% (EC_{50}). Thus, a low EC_{50} value represents a higher antioxidant activity in a sample (Hernández-Hernández et al., 2018b).This explains the negative correlations between the antioxidant activity (EC_{50}) and total phenols (-0.64) and, to a lesser extent, theobromine (-0.50). Negative and low correlations (≤- 0.4), although significant, were found between antioxidant activity and epicatechin derivatives and also between antioxidant activity and epigallocatechin. In accordance with our results, several studies have highlighted phenolic compounds as one of the most important contributors to the antioxidant activity of cocoa husk extracts (Arlorio et al., 2005; Manzano et al., 2017). Furthermore, theobromine contributes to antioxidant activity (Martínez et al., 2012; Okiyama et al., 2017).

High and positive Pearson's correlations (0.6–0.7) were found between epicatechin derivatives II, III and IV (Table 2). Also, total phenols were positively correlated with theobromine and most phenolic compounds (0.3–0.5), except catechin (-0.36). The total sugar content was positively correlated with theobromine and negatively correlated with catechin, although the coefficients were low (0.3 and -0.30, respectively). This negative correlation with catechin could be due to CBH's high sugar content and the effect of temperature, which enhances the epimerization and degradation of this phenol more rapidly than the degradation of epicatechin (Loncaric et al., 2017).

3.2. Spectral characteristics

The average spectra of intact cocoa beans and cocoa husk were displayed after their transformation with a second derivative spectral pre-treatment (Figure 1). Spectral patterns in the visible region were similar in shape (Figure 1), revealing absorption peaks from 474 to 648, which are related to chlorophyll contents (Morales-Sillero et al., 2011), and also to carotenoid and anthocyanin pigments (Strayer, 1995). In the NIR region, the average spectra of both types of products were quite similar, showing some discrepancies at specific areas in the range of 1.350-2.500 nm. Absorption bands between 1870 and 2002 nm are mainly associated with fat, and display higher values for the spectra of intact beans due to their higher fat content in relation to CBH. Absorption bands corresponding to sugars are found in the scientific literature at 1200 nm (Williams, 2001); total phenols were associated with 1349-1386, 1661-1718 and 2161-2258 nm (Hashimoto et al., 2018); while epicatechin absorption bands were reported at 1388, 1492, 1658, 1916, 2260 and 2324 nm; and those associated with theobromine at 1764, 2092 and 2228 nm (Álvarez et al., 2012). In Figure 1, it can be seen that absorption peaks in the bands 1716-1768 and 2280-2360 were more pronounced in the spectra of CBH, which could be related to the migration of phenols and theobromine from the cotyledon to the husk during the fermentation process (Hernández-Hernández et al., 2018b; Okiyama et al., 2017).

3.3. NIR calibration results

Twenty-two regression models were built to correlate the constituents of CBH with the VIS+NIR reflectance spectra of both CBH and intact cocoa beans in the range of

400–2500 nm. Five treatments of spectral derivatives were compared and the best performance was selected according to the lowest standard error of cross-validation (SECV) and the highest coefficient of determination for cross-validation (r^2).

Most of the best calibrations were developed with the first derivatives, in particular 1, 5, 5, 1, for total sugars, theobromine and epicatechin derivatives II, III and IV, and 1, 10, 10, 1 for epicatechin (Table 3). In the case of total phenols, catechin, derivative I, and epigallocatechin, the best calibrations were obtained with the second derivative 2, 5, 5, 1; while 2, 10, 5, 1 showed better results for estimating antioxidant activity.

The calibration statistics for total sugars, total phenols, catechin, and theobromine exhibited adequate accuracy as r^2 values were equal to or higher than 0.70. Coefficient of determination for total sugars was of 0.90 and the RPD was above 3, which indicated an excellent predictive ability (Nicolaï et al.,2007).The content in soluble sugars could be related to the presence of bioactive compounds, as previously reported in other agro-industrial by-products from olive oil and asparagus that had been pre-treated by a thermal or fermentation process. The use of temperature and enzymatic reactions are implied in the fermentation of cacao. These reactions enhance the production of bioactive compounds such as soluble sugars linked with phenols forming phenolic glycosides (Rubio-Senent et al., 2013) or antioxidant soluble fiber (Mrabet et al., 2017). In addition, fermentation has been shown to increase the formation of soluble neutral oligosaccharides and soluble pectin with functional and biological activities (Bermudez-Oria et al.,2017; Lama-Muñoz et al., 2012).

Theobromine and total phenols also showed some of the best calibration results, with r^2 values of 0.83 and 0.81, respectively; while the RPD values were between 2 and 2.5 (Table 3), which implies that coarse quantitative predictions are possible (Nicolaï et al., 2007). Considering the phenolic profile, only catechin showed good results, as the r^2 was 0.74 and the RPD was 1.98. This value for RPD implies that the model could be used to discriminate between low and high values (Nicolaï et al., 2007). The calibration models obtained for the rest of the extracted compounds showed poorer calibration

results ($r^2 \approx 0$). With the exception of epicatechin, these results are not surprising due to their low concentrations in the CBH samples (Table 1). According to Cozzolino (2015), a high degree of precision is generally found when applying near-infrared spectroscopy for the prediction of antioxidant compounds such as phenols, although one of the main disadvantages to this method is that it cannot separate individual phenolic compounds at low concentrations.

The calibration model for antioxidant activity showed an r^2 of 0.67 and an RPD of 1.74; thus this model can also be used to discriminate between low high values (Nicolaï et al., 2007). This result was expected, given the antioxidant activity attributed in the literature to CBH's total phenols and theobromine (Arlorio et al., 2005; Manzano et al., 2017; Martínez et al., 2012), and the relationships found between these parameters.

Similar calibration results were obtained from the spectra of intact cocoa bean (Table 4) compared to those obtained from the CBH spectra (Table 3). The second derivatives (2, 10, 5, 1 and 2, 5, 5, 1) provided the best calibration statistics for most of the constituents. Calibration statistics revealed similar predictive ability for total sugars compared to models developed out of husk spectra, which means that these compounds are well-predicted regardless of the type of spectra used (husk or intact bean spectra). The r^2 values for theobromine and total phenols obtained from intact beans decreased from 0.83 to 0.77 and from 0.81 to 0.71, respectively. The RPD also decreased, from 2.39 to 2.07 for theobromine, and from 2.28 to 1.85 for total phenols. However, coarse quantitative predictions are possible with the model developed for theobromine; while the model for total phenols allowed for discriminating between low and high values (Nicolaï et al., 2007). The calibration results for phenolic compounds from the intact bean spectra were similar to those obtained from the spectra of husk samples, and only the calibration model developed for catechin showed higher values for r^2 (0.62) and RPD (1.63), albeit lower than those obtained from the CBH spectra.

For all the studied parameters, the SECV values shown in Tables 3 and 4 were higher than the SEL values (Table 1), as $SECV^2 = SEL^2 + SE_{NIR}{}^2 + SE_{model}{}^2$, where

SE_{NIR} is the repeatability of the NIR method and SE_{model} represents the lack of fit of the calibration model (Fearn, 1986). The heterogeneity of the product, together with the use of coarse samples, always leads to higher errors in calibrations, as they increase the SE_{NIR}. In this sense, it was also observed that the SECV values for the calibrations obtained from the unground husk spectra were lower than those obtained from the intact cocoa bean spectra, as the reference values in the second data set are only referred to as a fraction (husk) of the scanned product (intact cocoa bean).

There were good correlations between reference and NIR predicted values for total sugars, total phenols and theobromine (Figure 2), with R^2 values equal to or higher than 0.9, regardless of the type of sample scanned (unground husk or intact bean). The highest values for total sugars (> 240 mg/g) corresponded to samples harvested in the 2014-2015 season, all from genotypes with different origins and grown in the experimental field in Tabasco. The differences in sugar concentration between different seasons may be due to the average temperature reached during the fermentation process, since higher temperatures increase the solubilization of sugars from the cell wall material, or synthesis of sugars in the fruit due to climatic conditions, or both. For theobromine contents, three groups of samples which basically corresponded to the harvest season were clearly identified. Thus, the CBH samples harvested in 2012-13 exhibited the lowest values (7-12 mg/g); while those from 2014-15 showed the highest contents (> 19 mg/g). The theobromine contents in the samples from the 2013-2014 season were intermediate, varying between 14 and 18 m/g.

There is no literature concerning the study of NIRS for the determination of bioactive compounds in CBH. Calibration performance could be compared to studies carried out with cocoa beans. In this sense, similar results were reported for the content in sugars by Krahmer *et al*. (2015) (r^2=0.82, and RPD=2.35) and Barbin et al. (2018) (R^2=0.94-0.99 and RPD=2.86-4.78, depending on pre-processing).The calibration models obtained in our study for total phenols are also in accordance with

the previous findings of Sunoj*et al*. (2016) (R^2=0.84-0.79, RPD=2.53-2.22, depending on the pre-processing) and Hashimoto *et al.*(2018) (R^2=0.89). A better calibration performance was reported by Kramer *et al.* (2015) (r^2=0.93, RPD=3.77), probably because only nine biological replicates were analyzed at different times of the fermentation process. However, the performance of NIR for the prediction of phenolic compounds has been scarcely studied. Only two works were found for the estimation of major compounds, i.e. epicatechin, which showed adequate predictive results, in contrast to our study. Álvarez et al. (2012) reported epicatechin values of R^2=0.96, SECV=0.18, and RPD= 2.3; while Kramer et al (2015) showed values of r^2=0.93 and RPD=3.69. The poorer results found in our work could be associated with the type of process undergone by the sample, as the cocoa beans were unfermented and sun-dried in the study of Alvarez et al. (2012), and non-fermented in the study of Krahmer *et al*. (2015). Regarding theobromine, Álvarez et al. (2012) indicated R^2=0.88, and RPD 2.5; Kramer *et al.* (2015) reported values of 0.79 for r^2 and 2.19 for RPD; whereas Hashimoto *et al* (2018) showed an R^2 of 0.77 -- all of them in accordance with those obtained in this work.

The regression coefficients for the best calibrations selected for total sugars, total phenols and theobromine from the spectra of intact cocoa beans and CBH are displayed in Figure 3. For total sugars, moderate/high coefficients were found for both studied products at wavelengths reported by Roggo et al. (2004) as associated with carbohydrates: i.e. combination C-H elongation/C-C elongation and C-O elongation at 2500 nm; combination C-H elongation/CH2 deformation at 2280-2330 nm; combination O-H elongation/ZOH deformation at 2100 nm; 1st overtone elongation at 1450 nm; 2nd overtone elongation at 1010-1030 nm; and 3rd overtone C-H elongation at 850-900 nm; with specific peaks around at 1620, 1660, 1680, 1790 and 1830 nm associated with sucrose. The coefficients displayed for the calibrations developed for total phenols showed several positive and negative peaks due to the use of spectral derivatives to develop the models. Among those sharp peaks, those cited by Hashimoto et al. (2018)

as associated with total phenol absorption were found at around 1349-1386, 1661-1718 and 2161-2258 nm. However, the coefficients obtained for theobromine coincided at 1764 nm with the absorption bands reported by Álvarez et al. (2012) but presented slight shifts at around 2094 and 2228 nm, which could also be attributed to the spectral derivatives used.

In view of the current findings, we can conclude that the use of NIRS technology for the classification and even for the analysis of CBH samples based on their contents in bioactive compounds (total sugars, total phenols and theobromine) is possible on an industrial scale. For this purpose, it would only be necessary to scan intact cocoa beans, thus avoiding the cumbersome process of separating the husk from the cotyledon. At a research level, measuring spectra directly on intact beans would also involve considerable time savings, since extraction is generally done by hand, particularly when rapid classification of samples is expected.

4. CONCLUSIONS

NIRS spectroscopy could be used with confidence in order to simultaneously predict and classify CBH constituents, such as total sugars, theobromine and total phenols, using spectra from the husk and even from intact beans. Therefore, this technology could be implemented as an economic, fast and environmentally-friendly alternative to conventional analysis methods currently used in food industry processes to extract bioactive compounds from this by-product of the cocoa industry.

5. ACKNOWLEDGEMENTS

This research was supported by the Spanish Ministry of Economy and Competitiveness and co-funded by the European Social Fund (ESF) (project AGL2016-79088R), the Spanish Ministry of Economy and Competitiveness Ramon y Cajal

Programme (RyC2012-10456), the National Institute of Forestry, Agricultural and Livestock Research (INIFAP-México),and the Mexican National Council of Science and Technology (CONACyT-Mexico).

6. CONFLICT OF INTEREST

The authors declare no conflict of interests.

7. REFERENCES

Adamson, G. E., Lazarus, S. A., Mitchell, A. E., Prior, R. L., Cao, G., Jacobs, P. H., … Schmitz, H. H. (1999). HPLC Method for the Quantification of Procyanidins in Cocoa and Chocolate Samples and Correlation to Total Antioxidant Capacity. https://doi.org/10.1021/jf990317m

Alander, J. T., Bochko, V., Martinkauppi, B., Saranwong, S., & Mantere, T. (2013). A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. *International Journal of Spectroscopy*, *2013*, 1–36. https://doi.org/10.1155/2013/341402

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., &Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (−)-epicatechin contents in unfermented and sun-dried beans of Criollococoa. *Journal of Near Infrared Spectroscopy*, 20(2), 307–315. https://doi.org/10.1255/jnirs.990

Amim, I., & Yee, C. L. (2006). Antioxidative Effects of Extracts of Cocoa Shell, Roselle Seeds and a Combination of Both Extracts on the Susceptibility of Cooked Beef to Lipid Oxidation. *Journal of Food Technology*. https://doi.org/doi=jftech.2006.10.15

Arlorio, M., Coisson, J. D., Restani, P., & Martelli, A. (2001). Characterization of Pectins and Some Secondary Compounds from Theobroma cacao Hulls. *Journal*

of Food Science, *66*(5), 653–656. https://doi.org/10.1111/j.1365-2621.2001.tb04616.x

Arlorio, M., Coïsson, J. D., Travaglia, F., Varsaldi, F., Miglio, G., Lombardi, G., & Martelli, A. (2005). Antioxidant and biological activity of phenolic pigments from Theobroma cacao hulls extracted with supercritical CO2. *Food Research International*, *38*(8–9), 1009–1014. https://doi.org/10.1016/j.foodres.2005.03.012

Azam, S., Hadi, N., Khan, N. U., & Hadi, S. M. (2003). Antioxidant and prooxidant properties of caffeine, theobromine and xanthine. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, *9*(9), BR325-R330.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, *43*(5), 772–777. https://doi.org/10.1366/0003702894202201

Bolliger, S., Zeng, Y., & Windhab, E. J. (1999). In-line measurement of tempered cocoa butter and chocolate by means of near-infrared spectroscopy. *Journal of the American Oil Chemists' Society*, *76*(6), 659–667. https://doi.org/10.1007/s11746-999-0157-5

Cozzolino, D. (2015). Infrared Spectroscopy as a Versatile Analytical Tool for the Quantitative Determination of Antioxidants in Agricultural Products, Foods and Plants. *Antioxidants*, *4*(3), 482–497. https://doi.org/10.3390/antiox4030482

Dong Young Kim, Hyen Joo Park, Hyung Hwan Park, Han Soo Kim, & Ik Boo Kwon. (1999). Manufacturing process of glucosyltransferase inhibitors from cacao bean husk, *37*(6).

Fayeulle, N., Vallverdu-Queralt, A., Meudec, E., Hue, C., Boulanger, R., Cheynier, V., Sommerer, N. (2018). Characterization of new flavan-3-ol derivatives in fermented cocoa beans. *Food Chemistry*, 259, 207-212. https://doi.org/10.1016/j.foodchem.2018.03.133

Gabriele, M., Gabriele, M., & Allan, G. (2006). No Title, (6).

Guthrie, J., Wedding, B., & Walsh, K. (1998). Robustness of NIR Calibrations for Soluble Solids in Intact Melon and Pineapple. *Journal of Near Infrared Spectroscopy*, *6*(1), 259–265. https://doi.org/10.1255/jnirs.145

Hashimoto, J. C., Lima, J. C., Celeghini, R. M. S., Nogueira, A. B., Efraim, P., Poppi, R. J., & Pallone, J. A. L. (2018). Quality Control of Commercial Cocoa Beans (Theobroma cacao L.) by Near-infrared Spectroscopy. *Food Analytical Methods*, *11*(5), 1510–1517. https://doi.org/10.1007/s12161-017-1137-2

Hernández-Hernández, C., Morales Sillero, A., Fernández-Bolaños, J., Bermúdez Oria, A., Azpeitia Morales, A., & Rodríguez-Gutiérrez, G. (2018). Cocoa bean husk: industrial source of antioxidant phenolic extract. *Journal of the Science of Food and Agriculture*. https://doi.org/10.1002/jsfa.9191

Hernández-Hernández, C., Viera-Alcaide, I., Morales-Sillero, A. M., Fernández-Bolaños, J., & Rodríguez-Gutiérrez, G. (2018). Bioactive compounds in Mexican genotypes of cocoa cotyledon and husk. *Food Chemistry*, *240*(August 2017), 831–839. https://doi.org/10.1016/j.foodchem.2017.08.018

Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. https://doi.org/10.1016/j.jfoodeng.2007.12.022

Lecumberri, E., Goya, L., Mateos, R., Alía, M., Ramos, S., Izquierdo-Pulido, M., & Bravo, L. (2007). A diet rich in dietary fiber from cocoa improves lipid profile and reduces malondialdehyde in hypercholesterolemic rats. *Nutrition*, *23*(4), 332–341. https://doi.org/10.1016/j.nut.2007.01.013

Lecumberri, E., Mateos, R., Izquierdo-Pulido, M., Rupérez, P., Goya, L., & Bravo, L. (2007). Dietary fibre composition, antioxidant capacity and physico-chemical properties of a fibre-rich product from cocoa (Theobroma cacao L.). *Food Chemistry*, *104*(3), 948–954. https://doi.org/10.1016/j.foodchem.2006.12.054

León, L., Garrido-Varo, A., & Downey, G. (2004). Parent and harvest year effects on

near-infrared reflectance spectroscopic analysis of olive (Olea europaea L.) fruit traits. *Journal of Agricultural and Food Chemistry*, *52*(16), 4957–4962. https://doi.org/10.1021/jf0496853

Manzano, P., Hernández, J., Quijano-Avilés, M., Barragán, A., Chóez-Guaranda, I., Viteri, R., & Valle, O. (2017). Polyphenols extracted from Theobroma cacao waste and its utility as antioxidant. *Emirates Journal of Food and Agriculture*, *29*(1), 45–50. https://doi.org/10.9755/ejfa.2016-04-388

Marten, G., Shenk, J., & Barton, F. (1989). Near infrared reflectance spectroscopy (NIRS): analysis of forage quality. *U.S. Department of Agriculture, Agriculture Handbook*, *643*, 1–110.

Martin-Cabrejas, M. a, Valiente, C., Esteban, R. M., Molla, E., & Waldron, K. (1994). Cocoa Hull: A Potential Source of Dietary Fibre. *J Sci Food Agric*, *66*(1991), 307–311. https://doi.org/10.1002/jsfa.2740660307

Martínez, R., Torres, P., Meneses, M. A., Figueroa, J. G., Pérez-Álvarez, J. A., & Viuda-Martos, M. (2012). Chemical, technological and in vitro antioxidant properties of cocoa (Theobroma cacao L.) co-products. *Food Research International*, *49*(1), 39–45. https://doi.org/10.1016/j.foodres.2012.08.005

Morales-Sillero, A., Fernández-Cabanás, V.-M., Casanova, L., Jiménez, M.-R., Suárez, M.-P., & Rallo, P. (2011). Feasibility of NIR spectroscopy for non-destructive characterization of table olive traits. *Journal of Food Engineering*, *107*(1). https://doi.org/10.1016/j.jfoodeng.2011.05.039

Mrabet, A., García-Borrego, A., Jiménez-Araujo, A., Fernández-Bolaños, J., Sindic, M., & Rodríguez-Gutiérrez, G. (2017). Phenolic extracts obtained from thermally treated secondary varieties of dates: Antimicrobial and antioxidant properties. *LWT - Food Science and Technology*, *79*, 416–422. https://doi.org/10.1016/j.lwt.2017.01.064

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality

by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, *46*(2), 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024

Okiyama, D. C. G., Navarro, S. L. B., & Rodrigues, C. E. C. (2017). Cocoa shell and its compounds: Applications in the food industry. *Trends in Food Science and Technology*, *63*, 103–112. https://doi.org/10.1016/j.tifs.2017.03.007

Ooshima, T., Osaka, Y., Sasaki, H., Osawa, K., Yasuda, H., & Matsumoto, M. (2000). Cariostatic activity of cacao mass extract. *Archives of Oral Biology*, *45*(9), 805–808. https://doi.org/10.1016/S0003-9969(00)00049-2

Osawa, K., Miyazaki, K., Shimura, S., Okuda, J., Matsumoto, M., & Ooshima, T. (2001). Identification of cariostatic substances in the cacao bean husk: Their anti-glucosyltransferase and antibacterial activities. *Journal of Dental Research*, *80*(11), 2000–2004. https://doi.org/10.1177/00220345010800111001

Peiris, K. H. S., Dull, G. G., Leffler, R. G., & Kays, S. J. (1998). Near-infrared spectrometric method for nondestructive determination of soluble solids content of peaches. *Journal of the American Society for Horticultural Science*. https://doi.org/10.21273/jashs.123.5.898

Roggo, Y., Duponchel, L., &Huvenne, J.P. (2004). Quality Evaluation of Sugar Beet *(Beta vulgaris)* by Near-Infrared Spectroscopy. Journal of *Agricultural and Food Chemistry,* 52 (5), 1055–1061. https://doi.org/10.1021/jf0347214

Serra Bonvehí, J., & Ventura Coll, F. (1999). Protein quality assessment in cocoa husk. *Food Research International*, *32*(3), 201–208. https://doi.org/10.1016/S0963-9969(99)00088-5

Singleton, V. L., Orthofer, R., & Lamuela-Raventós, R. M. (1998). Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent. *Methods in Enzymology*, *299*(1974), 152–178. https://doi.org/10.1016/S0076-6879(99)99017-1

Strayer, L. (1995). Biochemistry (4th ed.). New York: W.H.Freeman and Company/Worth Publishers.

Código de campo cambiado

Código de campo cambiado

Código de campo cambiado

Williams, P.C. (2001). Implementation of near-infrared technology. In P.C. Williams, & K.H.Norris (Eds.), *Near-infrared technology in the agricultural and food industries* St. Paul, Minnesota. AACC, Inc., pp. 145-169.

Witham, F.H., Blaydes, D.F. & Devlin, R.M. (1971). *Experiment in Plant Physiology*. Van Nostrand Reinhold Co., New York, NY, pp. 245.

Figure captions

Figure 1.- Second derivative of average spectra (Visible + NIR) of intact cocoa beans and cocoa beans husk samples.

Figure 2.- Cocoa husk reference (lab) data vs NIRS predicted values using a) unground cocoa husk spectra, and b) intact cocoa bean spectra.

Figure 3.- Regression coefficients for sugars (a), total phenols and theobromine of intact cocoa beans and cocoa beans husk samples.

Table 1. Descriptive statistical analysis of the parameters analyzed in cocoa beans husk (N = 80).

Constituent[a]	Range	Mean	SD	CV	SEL
Total sugars (mg/g)	12.35–337.09	72.05	75.29	104.50	11.18
Total phenols (mg/g)	6.95–80.04	23.52	13.37	56.83	3.51
Catechin (mg/g)	0.00–6.49	1.26	1.16	91.79	0.41
Epicatechin (mg/g)	4.40–46.42	16.43	6.92	42.15	4.34
Derivative I (mg/g)	0.00–17.54	0.48	2.02	423.67	0.12
Derivative II (mg/g)	0.00–3.14	0.22	0.49	219.66	0.11
Derivative III (mg/g)	0.00–2.28	0.12	0.35	294.01	0.06
Derivative IV (mg/g)	0.00–1.36	0.08	0.18	229.30	0.07
Epigalocatechin (mg/g)	0.00–1.36	0.14	0.23	167.53	0.07
Theobromine (mg/g)	7.39–54.59	19.70	10.74	54.51	1.51
Antioxidant activity (mg/mL)	5.59–76.88	26.84	15.78	58.78	2.82

SD: Standard deviation.

CV: Coefficient of variation (100*SD/Mean).

SEL: Standard error of laboratory.

a

All determinations were carried out in triplicate.

Table 2. Pearson's correlation between cocoa husk parameters for all samples.

Constituent	Total sugars	Total phenols	Catechin	Epicatechin	Derivative I	Derivative II	Derivative III	Derivative IV	Epigalocatechin	Theobromine
Total phenols	0.21ns									
Catechin	−0.30**	−0.36***								
Epicatechin	0.27*	0.30**	0.09ns							
Derivative I	0.00ns	0.41****	−0.17ns	0.29**						
Derivative II	−0.05ns	0.51****	−0.25*	0.04ns	0.08ns					
Derivative III	−0.04ns	0.41****	−0.12ns	0.05ns	0.05ns	0.72****				
Derivative IV	0.04ns	0.45****	0.04ns	0.12ns	0.05ns	0.61****	0.57****			
Epigalocatechin	−0.12ns	0.34**	−0.05ns	0.04ns	0.03ns	0.61****	0.37***	0.57****		
Theobromine	0.32**	0.49****	−0.63****	0.11ns	0.15ns	0.32**	0.16ns	0.10ns	0.26*	
Antiox. activity	−0.09ns	−0.64****	0.28*	−0.20ns	−0.22*	−0.37***	−0.27*	−0.24*	−0.31**	−0.50****

ns: non-significant.

*, **, *** and ****, significant at $P \leq 0.05$, 0.01, 0.001 and 0.0001.

Table 3. Calibration statistics for NIR models developed with the cocoa bean husk spectra.

Constituent	Derivative	N	Mean	SD	SEC	R^2	SECV	r^2	RPD
Total sugars	1,5,5,1	75	61.72	62.72	14.90	0.94	19.86	0.90	3.16
Total phenols	2,5,5,1	77	21.72	10.86	3.03	0.93	4.75	0.81	2.28
Catechin	2,5,5,1	80	1.18	1.08	0.43	0.84	0.55	0.74	1.98
Epicatechin	1,10,10,1	79	15.40	5.15	4.91	0.09	5.31	−0.06	0.97
Derivative I	2,5,5,1	80	0.18	0.30	0.26	0.23	0.27	0.16	1.09
Derivative II	1,5,5,1	77	0.11	0.19	0.17	0.15	0.18	0.08	1.05
Derivative III	1,5,5,1	78	0.05	0.09	0.09	0.09	0.10	−0.11	0.94
Derivative IV	1,5,5,1	78	0.05	0.08	0.07	0.05	0.08	0.00	0.99
Epigallocatechin	2,5,5,1	78	0.09	0.11	0.08	0.45	0.10	0.20	1.10
Theobromine	1,5,5,1	78	18.27	8.87	2.60	0.91	3.72	0.83	2.39
AntioxidantActivity	2,10,5,1	78	25.51	13.02	6.74	0.73	7.48	0.67	1.74

N: Number of samples used for calibration.

Mean: Mean of the calibration series.

SD: Standard deviation.

SEC: Standard calibration error.

R^2: Determination coefficient for calibration.

SECV: Standard cross calibration error.

r^2: Determination coefficient for cross-validation.

RPD: Ratio SD/SECV.

Table 4. Calibration statistics for NIR models developed with the intact cocoa bean spectra.

Constituent	Derivative	N	Mean	SD	SEC	R^2	SECV	r^2	RPD
Total sugars	2,10,5,1	77	70.35	76.31	19.64	0.93	24.07	0.90	3.17
Total phenols	2,10,5,1	74	22.01	11.25	3.64	0.90	6.09	0.71	1.85
Catechin	2,10,5,1	76	1.15	1.06	0.40	0.86	0.65	0.62	1.63
Epicatechin	1,10,5,1	75	15.23	5.16	4.86	0.11	5.24	−0.04	0.98
Derivative I	2,5,5,1	76	0.18	0.30	0.24	0.39	0.27	0.18	1.11
Derivative II	2,10,5,1	75	0.11	0.19	0.13	0.49	0.17	0.20	1.12
Derivative III	2,5,5,1	71	0.03	0.05	0.05	0.09	0.05	−0.03	0.97
Derivative IV	2,10,5,1	76	0.05	0.08	0.07	0.06	0.08	−0.05	0.97
Epigallocatechin	1,5,5,1	72	0.08	0.09	0.08	0.10	0.09	−0.02	0.97
Theobromine	2,5,5,1	75	18.70	9.43	3.35	0.87	4.55	0.77	2.07
AntioxidantActivity	2,5,5,1	77	25.09	13.16	7.08	0.71	8.90	0.54	1.48

N: Number of samples used for calibration.

Mean: Mean of the calibration series.

SD: Standard deviation.

SEC: Standard calibration error.

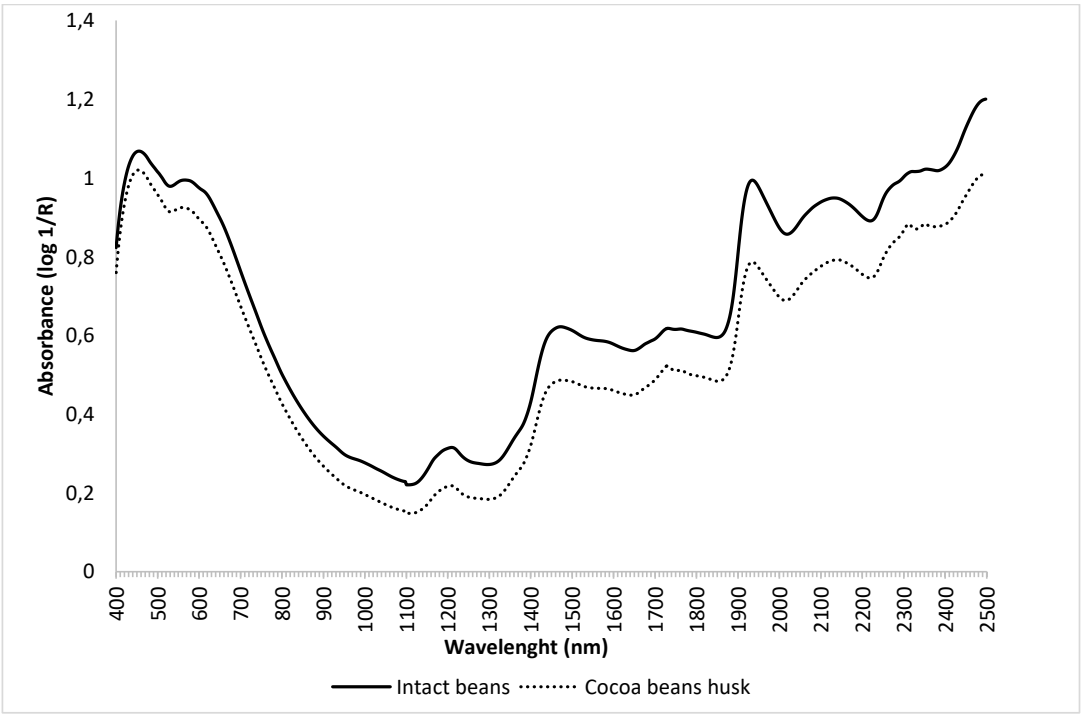
R^2: Determination coefficient for calibration.

SECV: Standard cross calibration error.

r^2: Determination coefficient for cross-validation.

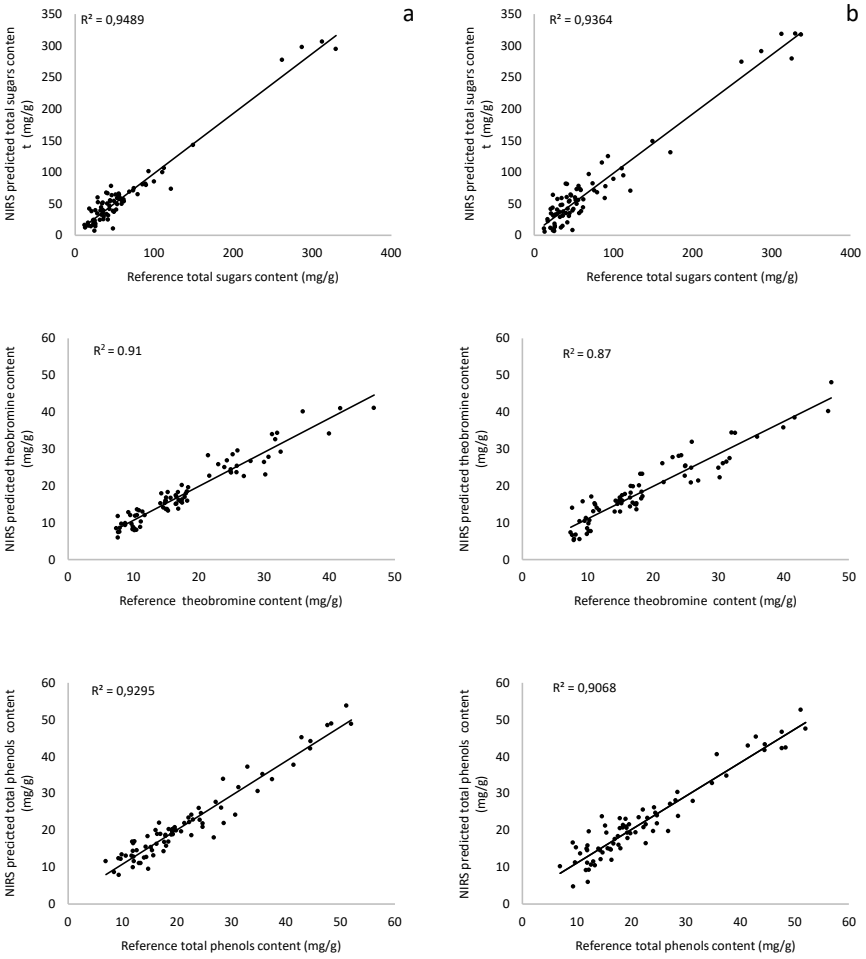
RPD: Ratio SD/SECV.

Figure 1


1,4
1,2
1
0,8
0,6
0,4
0,2
0
Absorbance (log 1/R)
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
Wavelenght (nm)
Intact beans
Cocoa beans husk

28 Figure 2

29



30

31

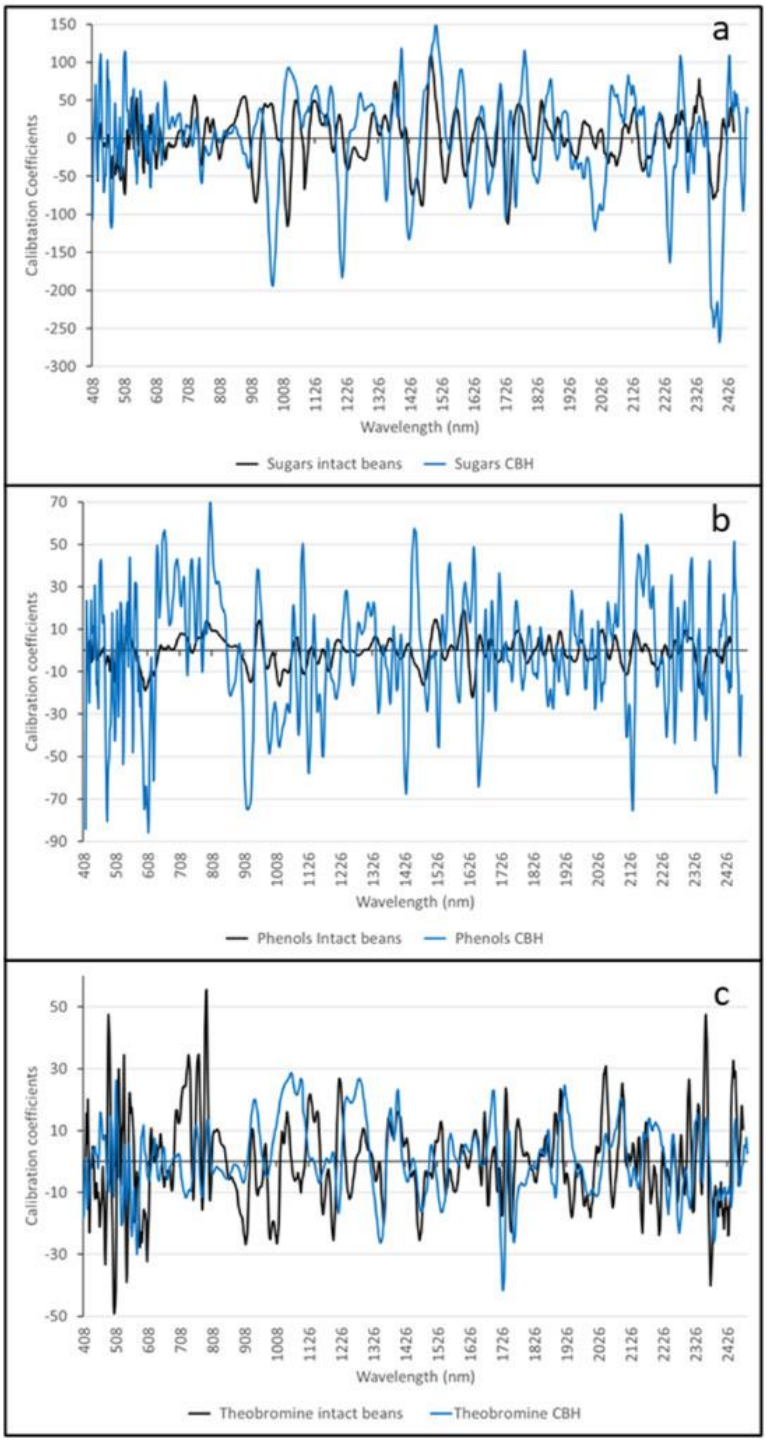
32

33

34

35

Figura 3


a
Calibtation Coefficients
Wavelength (nm)
Sugars intact beans
Sugars CBH
b
Calibration coefficients
Wavelength (nm)
Phenols Intact beans
Phenols CBH
c
Calibration coefficients
Wavelength (nm)
Theobromine intact beans
Theobromine CBH

39

40